



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2014

3'-UTR Poly(T/U) tract deletions and altered expression of EWSR1 are a hallmark of mismatch repair-deficient cancers

Kishore, Shivendra ; Piscuoglio, Salvatore ; Kovac, Michal B ; Gylling, Annette ; Wenzel, Friedel ;
Trapani, Francesca ; Altermatt, Hans Joerg ; Mele, Valentina ; Marra, Giancarlo ; Peltomäki, Päivi ;
Terracciano, Luigi ; Zavolan, Mihaela ; Heinemann, Karl

Abstract: The genome-wide accumulation of DNA replication errors known as microsatellite instability (MSI) is the hallmark lesion of DNA mismatch repair (MMR)-deficient cancers. Although testing for MSI is widely used to guide clinical management, the contribution of MSI at distinct genic loci to the phenotype remains largely unexplored. Here, we report that a mononucleotide (T/U)₁₆ tract located in the 3' untranslated region (3'-UTR) of the Ewing sarcoma breakpoint region 1 (EWSR1) gene is a novel MSI target locus that shows perfect sensitivity and specificity in detecting mismatch repair-deficient cancers in two independent populations. We further found a striking relocalization of the EWSR1 protein from nucleus to cytoplasm in MMR-deficient cancers and that the nonprotein-coding MSI target locus itself has a modulatory effect on EWSR1 gene expression through alternative 3' end processing of the EWSR1 gene. Our results point to a MSI target gene-specific effect in MMR-deficient cancers. *Cancer Res*; 74(1); 224-34. ©2013 AACR.

DOI: <https://doi.org/10.1158/0008-5472.CAN-13-2100>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-89778>

Journal Article

Accepted Version

Originally published at:

Kishore, Shivendra; Piscuoglio, Salvatore; Kovac, Michal B; Gylling, Annette; Wenzel, Friedel; Trapani, Francesca; Altermatt, Hans Joerg; Mele, Valentina; Marra, Giancarlo; Peltomäki, Päivi; Terracciano, Luigi; Zavolan, Mihaela; Heinemann, Karl (2014). 3'-UTR Poly(T/U) tract deletions and altered expression of EWSR1 are a hallmark of mismatch repair-deficient cancers. *Cancer Research*, 74(1):224-234.

DOI: <https://doi.org/10.1158/0008-5472.CAN-13-2100>



Cancer Research

3'UTR poly(T/U) tract deletions and altered expression of EWSR1 are a hallmark of mismatch repair deficient cancers

Shivendra Kishore, Salvatore Piscuoglio, Michal Kovac, et al.

Cancer Res Published OnlineFirst October 24, 2013.

Updated version	Access the most recent version of this article at: doi: 10.1158/0008-5472.CAN-13-2100
Supplementary Material	Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2013/10/24/0008-5472.CAN-13-2100.DC1.html
Author Manuscript	Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org .
Permissions	To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org .

3'UTR poly(T/U) tract deletions and altered expression of *EWSR1* are a hallmark of mismatch repair deficient cancers

Shivendra Kishore^{1*}, Salvatore Piscuoglio^{2*}, Michal Kovac^{2*}, Annette Gylling³, Friedel Wenzel², Francesca Trapani^{4,8}, Hans Joerg Altermatt⁶, Valentina Mele^{4,5}, Giancarlo Marra⁷, Päivi Peltomäki³, Luigi Terracciano⁴, Mihaela Zavolan^{1\$} and Karl Heinimann^{2\$}.

¹Biozentrum, University of Basel, Basel, Switzerland.

²Research Group Human Genetics, Department of Biomedicine and Division of Medical Genetics, University of Basel, Basel, Switzerland.

³Department of Medical Genetics, Haartman Institute, University of Helsinki, Helsinki, Finland.

⁴Institute of Pathology, University Hospital of Basel, Basel, Switzerland.

⁵Institute for Surgical Research and Hospital Management and Department of Biomedicine, University of Basel, Basel, Switzerland.

⁶Pathologie Laenggasse Bern, Bern, Switzerland

⁷Institute of Molecular Cancer Research, University of Zurich, Zurich, Switzerland.

⁸Paterson Institute for Cancer Research, University of Manchester, Manchester, UK.

** These authors equally contributed to the work*

^{\$}Corresponding authors:

Mihaela Zavolan
Biozentrum
University of Basel
Klingelbergstrasse 50 / 70
4056 Basel, Switzerland
Phone: +41 61 267 15 77
E-mail: mihaela.zavolan@unibas.ch

and

Karl Heinimann
Research Group Human Genetics
Department of Biomedicine, University of Basel
Mattenstrasse 28
4058 Basel, Switzerland
Phone: +41 61 265 36 20
Fax: +41 61 265 36 21
E-mail: karl.heinimann@unibas.ch

Running title: EWS16T a novel genetic marker for microsatellite instability

Keywords: DNA mismatch repair, Microsatellite instability, Lynch Syndrome, colorectal cancer

Figure 6; Tables: 1

Words: 4530

The authors disclose no potential conflicts of interest

Abstract

The genome-wide accumulation of DNA replication errors known as microsatellite instability (MSI) is the hallmark lesion of DNA mismatch repair (MMR) deficient cancers. Although testing for MSI is widely used to guide clinical management, the contribution of MSI at distinct genic loci to the phenotype remains largely unexplored. Here we report that a mononucleotide (T/U)₁₆ tract located in the 3' untranslated region (3'UTR) of the Ewing sarcoma breakpoint region 1 (EWSR1) gene is a novel MSI target locus that shows perfect sensitivity and specificity in detecting mismatch-repair deficient cancers in two independent populations. We further found a striking re-localization of the EWSR1 protein from nucleus to cytoplasm in MMR-deficient cancers, and that the non-protein coding MSI target locus itself has a modulatory effect on EWSR1 gene expression through alternative 3' end processing of the EWSR1 gene. Our results point to a MSI target gene-specific effect in MMR-deficient cancers.

Précis

This study identifies a novel genetic locus that is fully informative and accurate in detecting mismatch-repair deficient cancers, with major implications for routine daily practice in the clinic.

Introduction

With an estimated population incidence of about 1 in 370, Lynch syndrome (LS, formerly known as hereditary non-polyposis colorectal cancer, HNPCC) represents the most common, autosomal dominantly inherited cancer predisposition world-wide(1), caused by germline mutations in DNA mismatch repair (MMR) genes (MLH1, MSH2, MSH6 and PMS2). Mutation carriers are at an increased risk to develop colorectal cancer (CRC) as well as a distinct spectrum of extracolonic cancers (predominantly of the endometrium, ovaries and stomach) at a young age(2).

At the cellular level, biallelic inactivation of the MMR system leads to genome-wide accumulation of DNA replication errors at specific repetitive nucleotide sequences, a condition termed microsatellite instability (MSI). MSI is the hallmark lesion of MMR deficient cancers(3), but is also observed in 10 to 20% of the sporadic colorectal, gastric and endometrial cancers. The tumour's MSI status is increasingly used to guide clinical management(3, 4), because genome-wide gene expression data from (sporadic) microsatellite stable (MSS) and unstable (MSI) CRCs demonstrated that tumor development in MMR-deficient cancers follows distinct pathogenetic paths(5).

A large number of distinct genic loci affected by MSI have been described, consisting mainly of mono- and dinucleotide repeats within the 5'UTR (e.g. NR-27), introns (BAT-26) and the 3'UTR (CAT25) of specific genes(6). The functional significance of MSI at these non-coding repeat loci and how they may contribute to the pathogenic process is, however, largely unknown(7). Here we report a novel target gene locus, EWS16T, consisting of a monomorphic polythymine (16T) tract within the 3'UTR of the Ewing sarcoma gene (*EWSR1*). We assessed this locus in two independent populations and found that EWS16T discriminates MMR-deficient from MMR-proficient cancers, be they sporadic or hereditary, with high sensitivity and specificity. Despite its location in a non-coding region of the *EWSR1* gene, EWS16T contractions are associated with changes in *EWSR1* expression and subcellular localization. Our findings thus directly implicate the RNA/DNA-binding Ewing sarcoma protein, better known for its fused variants in Ewing sarcoma, in MSI-associated colorectal tumorigenesis.

Material and Methods

DNA and RNA Isolation

To isolate genomic DNA and total RNA from cell lines and fresh/frozen tumour tissue samples Qiagen QIAamp DNA/RNA Mini Kit (Qiagen, Hombrechtikon, Switzerland) and for formalin fixed paraffin embedded (FFPE) tumour samples RecoverAll™ Total Nucleic Acid Isolation Kit (Ambion, Invitrogen, Carlsbad, CA, USA) were used according to the manufacturers' guidelines.

Analysis of Microsatellite Instability (MSI)

Microsatellite instability was assessed on two independent cohorts of patients encompassing 85 Lynch syndrome related cancers (78 CRCs and 7 endometrial carcinomas), 113 sporadic CRCs, including 14 cases with MLH1 promoter hypermethylation and 12 MMR-proficient, MSI-low cancers, and a Finnish cohort including 8 Lynch syndrome gastric related cancers, 50 sporadic CRCs and 64 sporadic gastric cancers, all of which were compared to matched tumor free mucosa. Based on the recommendations of the National Cancer Institute workshop on MSI, a panel of microsatellite loci (BAT25, BAT26, D2S123, D5S346, D17S250) and two additional microsatellite markers (BAT40, MYCL1) were used to determine MSI status(8). The 3'UTR poly T(16) tract (EWS16T) of the EWSR1 gene (Ewing sarcoma breakpoint region 1; RefSeq: NM_005243) was amplified by PCR with the following primers: 5'- AATGTTTCATGGTTGTGATGT-3' (forward FAM-labelled) and 5'- GAAGGATGACTCTTTATAA-3' (reverse). PCR products were analysed on an ABI 310 Genetic Analyzer with GeneScan Analysis V 3.1 (PE Applied Biosystems, Foster City, CA) and Genotyper 2.0 (PE Applied Biosystems, Foster City, CA) software. Fragment analysis of PCR products allowed determination of novel alleles (expansions or contractions) within the repetitive tract of a given marker. The observed expansions or contractions in EWS16T were verified using a second set of primers covering the locus (Forward primer: 5'- GCATGCTCAGTATCATTGTGG-3'; Reverse primer: 5'-AGGCCGAGAAGGATGACTCT-3') for sequencing analysis of selected samples. Sequencing reactions using the Big Dye terminator chemistry (Applied Biosystems, Foster City, CA) were performed according to the manufacturer's protocol.

Relative expression of EWSR1 by qPCR

EWSR1 mRNA expression on fresh frozen tissues (8 Lynch syndrome related CRCs and 5 sporadic CRCs both matched with their tumour free mucosa) was assessed using the TaqMan® Probe-Based Gene Expression Analysis (Applied Biosystems, Foster City, CA), and the EWSR1 probe Hs01580532_g1 (Applied Biosystems, Foster City, CA). The measurements were normalized using the HPRT1 probes Hs01003270_g1 (Applied Biosystems, Foster City, CA)(9, 10), and the fold-changes in gene expression were calculated using the standard $\Delta\Delta C_t$ method(11). All retrotranscriptase reactions, including no-template controls, were run on an Applied Biosystem 7900HT thermocycler. Each sample was tested in triplicate unless specified otherwise.

Immunohistochemistry

Several cohorts of patients were studied by immunohistochemical analysis of EWS. Briefly, the tissue samples of the following cohorts of patients were analysed: 37 colon adenomas (9 of which Lynch syndrome related), 19 CRCs (10 of which Lynch syndrome related) and a tissue microarray(12). Patient data including complete follow-up were obtained by retrospective analysis of medical records, regional tumour registries and/or treating physicians. Tissue samples were obtained by surgical or endoscopic excision. Tissue sections of 4 μ m sections of paraffin embedded tissue were immunostained for primary antibody against EWS (Abcam clone 84389 dilution 1:800). Staining was carried out as previously described(13). Immunoreactivity was scored semi-quantitatively by evaluating the number of positive tumour cells over the total number of tumour cells. Nuclear immunoreactivity scores were assigned using 5% intervals and ranged from 0% to 100%. Regarding cytoplasmic expression, the staining intensity was scored as described by Allred et al.(14). All samples were examined independently by three different pathologists (S.P., F.T. and L.T.), blinded to clinicopathological and molecular genetic information.

Cell lines

Human cell lines included in this study were purchased from American Type Culture Collection (ATCC, Rockville, MD) and authenticated by ATCC by short tandem repeat (STR) profiling and used for functional studies within 6 months after thaw from liquid nitrogen tank. All cell lines were maintained at 37°C with 5% CO₂.

Five colorectal cancer cell lines were used for this study: three mismatch repair (MMR) deficient cell lines (HCT116, LoVo, HCT15) and two MMR proficient ones (SW480 and HT29). HCT116 and HCT15 cells were cultured in RPMI 1640 (Invitrogen Basel, Switzerland) supplemented with 10% fetal bovine serum (FBS), 1% Kanamycin sulphate, 1% GlutaMAX-I, 1% Sodium Pyruvate, 1% non Essential Amino Acids (NEAA), 1% HEPES (all from Invitrogen Basel, Switzerland) and 0.1% 2-mercapto-ethanol (Sigma-Aldrich Basel, Switzerland). LoVo cells were cultured in Ham's F12 media supplemented with 10% FBS. HT29 cells were grown in McCoy's 5A Medium (Invitrogen Basel, Switzerland) with 10% FBS, Kanamycin sulphate and GlutaMAX-I (all from Invitrogen Basel, Switzerland). SW480 cells were cultured in L-15 Medium (Sigma-Aldrich Basel, Switzerland) with 10% FBS, 1% GlutaMAX-I and 1% Kanamycin sulphate (all from Invitrogen Basel, Switzerland).

HeLa cells at earlier passages were cultured in DMEM with GlutaMAX (Invitrogen) supplemented with 10% FBS.

Plasmids and antibodies

Synthetic poly(A) site of Renilla luciferase gene in dual luciferase psiCHECK-2 vector was mutated using primers psi-ck2polyAmutF: 5'-GCGGCCGCTGGCCGCAGCTAAATATCTTTATTTTCA-3' and psi-ck2polyAmutR: 5'-TGAAAATAAAGATATTTAGCTGCGGCCAGCGGCCGC-3' using QuikChange™ Site-Directed Mutagenesis Kit as per manufacturer's instruction to generate psiCHECK-2-Synthetic polyA mutant (psiCHECK-2-SPAm). EWSR1 3' UTR was PCR amplified from HEK293 genomic DNA using primers EWSR1_xhoIF: 5'-

CCGACTCGAGCGGCCCTACTAGATGCAGAG-3' and EWSR1_notIR: 5'-ATAAGAATGCGGCCGCGAACCAACCGTTTACCTGGA-3'. The PCR amplicons were cloned into pGEM-T Easy and subsequently subcloned into psiCHECK-2-SPAm using XhoI and NotI restriction sites to generate EWSR1-wt-3'UTR/psiCHECK-2-SPAm reporter. Mutant constructs with U deletions (EWSR1-2/3/4/5/6 del-3'UTR/psiCHECK-2-SPAm) and insertions (EWSR1-2/4/ins-3'UTR/psiCHECK-2-SPAm) were introduced using standard overlap extension PCRs.

Antibodies for Western blots against NF90, SFPQ and HuR were obtained from SantaCruz Biotech and Antibody against EWSR1 for Western blots and Immunoprecipitation were obtained from Abcam.

In vitro transcription

In vitro transcription for pull down assay using S1 aptamer was performed using T7 RiboMAX™ Express Large Scale RNA Production System (Promega) as per manufacturer's instructions. Region flanking the wt- and deleted poly T/U tracks in the 3'UTR of EWSR1 constructs was amplified using primers T7_EWSR1_IVT_Cf: 5'-GCTTCTAATACGACTCACT ATAGGGAGAAATGGGAACCCCTTGTGAG-3' and EWSR1_IVT_Cr: 5'-GAACAGAGGCCCGA GAAGGAT-3' to introduce T7 promoter sequences at the 5' end of the amplicons. S1 aptamer sequence was introduced at the 3' end using another round of PCR using T7_EWSR1_IVT_Cf: 5'-GCTTCTAATACGACTCACTATAGGGAGAAATGGGAACCCCTTGT GAG-3' and EWSR1_S1apt_IVT_Cr: 5'-CATGGCCCCGCCCCGCGACTATCTTACGCACTTG CATGATTC TGGTCGGTCCCATGGATCCGAACAGAGGCCGAGAAGGAT-3'.

Protein Pull-Down Assay

For each sample, 100ul of the MyOne Streptavidin Dynal beads (Invitrogen) was washed twice with one bead volume of solution A (DEPC-treated 0.1 M NaOH, DEPC-treated 0.05 M NaCl) and once with one bead volume of Solution B (DEPC-treated 0.1 M NaCl) and once with RNA binding buffer. The beads were resuspended in one bead volume of RNA binding

buffer (100mM NaCl, 50mM Hepes 7.5, 0.5% NP-40 and 10mM MgCl₂) with 100ug of *in vitro* transcribed RNA with S1 aptamer sequence and incubated at 10°C for 40 minutes in a thermomixer with intermittent shaking. The beads were washed twice with one volume of RNA wash buffer prior to incubation with the lysate. HEK293 cell pellet from 15cm² dish was lysed in 3ml native lysis buffer (25mM Hepes-KOH pH 7.5, 100mM KCl, 0.5% NP-40, 5mM MgCl₂, 0.5mM DTT, protease inhibitor cocktail, 1mM NaF, 1 mM Na₄VO₄ and 300U of RNasin) for 15 minutes on ice. The lysate was subsequently gently sonicated and centrifuged to remove any cell debris. 200ug of E. coli tRNA was additionally added to prevent non-specific binding of proteins to the beads. 1 ml of the lysate was added to the beads coupled to S1 aptamer RNA and also to the beads alone for no RNA control. The mixture was incubated at 4°C on rotation wheel. After 1 hour of incubation, the beads were washed thrice with native lysis buffer. The bound proteins were eluted with 100ul native lysis buffer supplemented with 25mM Biotin for 30-45 minutes at 10°C on a thermomixer with intermittent shaking. 900ul of 100% ethanol was added to the eluate and incubated at -80°C for 2 hours followed by centrifugation to precipitate the eluted proteins. The pellet was air dried and dissolved in 35ul of (SDS loading dye). Prior to loading on the NuVex gradient gels, the samples were heated at 90°C for 5 minutes. After SDS-PAGE electrophoresis, the gel was stained with colloidal blue and bands of interest were excised and sent for mass spectrophotometry.

Poly(A) site selection assay

HeLa cells were transfected with psiCHECK-2-SPAm constructs for 24 hours. Total RNA was isolated from the HeLa cells using TriReagent (Sigma) followed by DNase I (Promega) treatment according to the manufacturer's protocol. Reverse transcription was done with oligo d(T)₁₈ primers for 1 hour. For poly(A) site selection assay, multiplexed polymerase chain (PCR) reaction was set up using a single forward primer specific to the psiCHECK-2 vector (psiCHECK-2-SeqFor: 5'-ATGAAATGGGTAAGTACA-3') and two 4 nucleotide terminally anchored oligo d(T)₁₆ reverse primers specific to the two isoform variants of EWSR1 (EWSpolyAproxR: 5'-TTTTTTTTTTTTTTTTTACCA-3' and EWSpolyAdistR: 5'-TTTTTTTTTTTTTTTTTGACT-3' respectively) to detect only the isoforms generated specifically from the psiCHECK-2-SPAm constructs. PCR was run for 28 cycles and the products were

separated on a 2 % Agarose gel. The bands were quantified using the ImageJ software (<http://rsbweb.nih.gov/ij/>).

Luciferase assays

HeLa cells were seeded in a 48 well plate one day prior to transfection. 0.2ug of plasmids (psiCHECK-2-SPAm constructs) were transfected with Lipofectamine 2000 (Invitrogen) for 24 hours. Luciferase assays were done on the transfected cells using The Dual-Luciferase® Reporter Assay System (Promega). Both transfections and luciferase assays were done according to the manufacturer's protocol.

siRNA transfections

Control-siRNAs and siRNAs against hnRNPC, HuR, NF90 and SFPQ and were obtained from Santa Cruz Biotechnology. HeLa cells were reverse transfected with siRNA oligos using RNAiMAX (Invitrogen). After 48 hours the cells were transfected with EWSR1-wt-3'UTR- and EWSR1-6del-3'UTR-psiCHECK-2-SPAm-constructs for another 24 hours. The cells were subsequently harvested and split into two aliquots. One aliquot was used to assess the knockdown efficiency of siRNA using Western blot, while the other was used for RNA isolation and subsequent poly(A) site selection assay.

Statistical Analyses

For statistical analysis, the chi-square test (χ^2 test) and Fisher's exact test for nonparametric variables and Student's t-test for parametric variables were used, with all probabilities reported as 2-tailed, considering a $P < 0.05$ to be statistically significant. Calculations were performed using the software program SPSS 17.0 (IBM Corporation, Somers, NY 10589).

Ethical approval

The study is part of the so-called "Basler Studie über familiaere Tumorkrankheiten", Ref.Nr. EK258/05 and has been approved by the "Ethikkommission beider Basel". Furthermore, written informed consent was obtained from all Lynch syndrome patients as well as from the sporadic patients.

Results

The initial investigation of a primary (vimentin- and CD99-positive) extra-osseous Ewing sarcoma (ES) from a *MSH6* mutation carrier (c.3696dupT) previously affected by colon cancer revealed that the tumor displayed the typical hallmarks of LS, i.e. MSI at mononucleotide markers and specific loss of *MSH6* expression (Supplementary Fig. S1), but none of the molecular features commonly associated with ES, i.e. chromosomal translocations involving the *EWSR1* gene locus on 22q12 (Supplementary Fig. S2). Reasoning that the *EWSR1* gene is likely involved in the development of this Ewing sarcoma tumour and in view of the tumour's MSI-high phenotype we focused on a mononucleotide tract in the 3' UTR of the *EWSR1* gene consisting of 16 thymines (EWS16T; c.*318_*333) which we thought would be a likely MSI target. Strikingly, both of the patient's tumors, the Ewing sarcoma as well as the colon cancer, were found to carry somatic contractions/deletions of 4 and 5 thymines, respectively. This prompted us to assess the prevalence of somatic alterations at EWS16T in MMR-deficient cancers in general. In contrast to Wheeler et al.(15), who reported a deletion variant (c.*331_*333delTTT; dbSNP: rs76631619) in James D. Watson's genome, we found the EWS16T locus to be quasi-monomorphic in 370 (99.7%) out of 371 constitutional (i.e. tumor-free) DNA samples tested.

EWS16T: a novel target gene locus in Lynch Syndrome-related cancers.

To determine whether EWS16T is indeed a novel target gene locus in Lynch syndrome (LS) we analysed 85 matched tumor-free tissue and cancer samples (78 colorectal (CRC) and 7 endometrial) from 78 Swiss LS patients with confirmed MMR germline mutation (58 *MLH1*, 20 *MSH2*). We further investigated 14 sporadic MMR-deficient CRCs with *MLH1* promoter hypermethylation as well as 86 sporadic MMR-proficient CRCs. Assessment of EWS16T tract length by capillary electrophoresis of fluorescently labelled PCR products revealed that all MMR-deficient cancers but none of the MMR-proficient ones displayed novel alleles, i.e. contractions or expansions at the EWS16T tract (Fig. 1). These findings were further confirmed independently in a Finnish cohort of 122 patients. In this patient cohort as well all of the 29 MMR-deficient (12 CRC, 17 gastric cancers) but none of the 93 MMR-proficient (38 CRC and 55 gastric) cancers showed EWS16T tract instability. The majority (72.7%) of somatic alterations consisted of contractions/deletions of 4 or more base pairs

(Supplementary Fig. S3). Consistently, we found that all MMR-deficient cell lines investigated (LoVo, HCT15, HCT116) carried solely mutated EWS16T alleles (contractions) without evidence for a wild-type tract allele. In contrast, MMR-proficient cell lines (HT29, SW480) as well as 37 microsatellite-stable colon adenomas and 12 MMR-proficient, MSI-low CRCs from Swiss patients had a stable EWS16T locus (Table 1). Thus, the EWS16T tract represents a novel, quasi-monomorphic MSI target locus identifying hereditary and sporadic MMR-deficient cancers with 100% sensitivity (95% CI 97-100) and specificity (95% CI 98-100).

EWSR1 poly(T/U) tract deletions promote distal poly(A) site usage.

Fusion of *EWSR1* with ETS family transcription factors as a result of chromosomal translocations has long been implicated in the development of Ewing sarcoma(16). It is still debated to what extent tumorigenesis is the consequence of the fusion protein as opposed to the loss of function of the normal EWSR1 protein. Because somatic EWS16T tract alterations were exclusively present in MMR-deficient cancers, whether hereditary or sporadic, and occurred in all types of cancer investigated (colorectal, gastric and endometrial) we wondered about their possible functional role(s) in MMR-related carcinogenesis. The poly(T/U) tract deletions occur in a region which encodes the 3'UTR of *EWSR1* and therefore do not alter the coding sequence of the *EWSR1* gene. Nonetheless, 3'UTRs contain sequence elements that are important for the post-transcriptional regulation of protein levels. Furthermore, it has been recently demonstrated that changes in 3'UTR length through alternative polyadenylation activates oncogenes(17). We thus set out to characterize the effect of EWS16T tract deletions on *EWSR1* expression levels. A recently generated catalogue of polyadenylation sites identified through 3' end sequencing from human embryonic kidney (HEK) 293 cells(18) shows that *EWSR1* undergoes alternative polyadenylation, generating two transcript forms that differ in the length of their 3'UTRs (Supplementary Fig. S4). The ESW16T tract deletions occur very close to the distal poly(A) signal and may thus result in changes in 3' end processing factor assembly, thereby altering the poly(A) site selection. To investigate this possibility, we cloned the 3'UTR of *EWSR1* downstream of the Renilla luciferase-coding region in a psiCHECK-2 mutant vector in which the synthetic poly(A) site was mutated (psiCHECK-2-SPAm). This construct thus allowed only the usage of the poly(A) signal from

the cloned 3'UTR of *EWSR1*. We then generated variant constructs containing poly(T/U) tracts of variable lengths through deletion mutation. With primers that simultaneously detect both the short and the long 3'UTR isoforms in a multiplexed semi-quantitative PCR, we found that deletions in the EWS16T tract promoted the usage of the distal poly(A) site (Figure 2a-b). We further investigated the MMR-proficient and MMR-deficient colon cancer cell lines and found that, consistent with our findings in the heterologous system, MMR deficiency is associated with higher expression of the longer *EWSR1* isoform (Figure 2c). These results indicate that the EWS16T tract deletions indeed alter poly(A) site selection.

SFPQ but not HuR or NF90 mediates distal poly(A) site usage in *EWSR1* pre-mRNAs with EWS16T tract deletions.

To determine the factors involved in *EWSR1* poly(A) site selection, we used S1 aptamer-tagged, *in vitro* transcribed wildtype 3'UTR and a 3'UTR variants with 6U deletions in the EWS16T region to pull down the proteins that associated with these RNAs (Supplementary Fig. S5). In three independent experiments we reproducibly identified a set of A/U-rich element binding proteins that associate with these constructs (Figure 3A and 3B). Interestingly, we found that nuclear factor 45/90/110 (NF45/90/110), heterogeneous nuclear ribonucleoprotein C (hnRNPC) and human antigen R (HuR) associate with the wildtype but not with the mutant 3'UTR. NF45 and NF90 have been previously shown to be part of a heterodimeric complex, nuclear factor of activated T-cells (NFAT), which is required for T-cell expression of interleukin 2, with NF110 being the larger isoform of NF90. NF90 has been shown to regulate mRNA stability and redistribution of nuclear mRNAs in the cytoplasm(19). Conversely, the EWS16T mutant preferentially associated with the SFPQ/NONO heterodimer, which is an essential pre-mRNA splicing factor that couples splicing with polyadenylation as a component of a small nuclear ribonucleoprotein (snRNP)-free complex with SNRPA/U1A(20). Strikingly, NONO has recently been implicated in the cellular response to DNA double-strand breaks [PMID: 22941645].

To determine which of the RNA binding proteins (RBP) identified above influenced the poly(A) site selection, we knocked down their expression individually with siRNAs (Figure 4a) and

assessed the poly(A) site usage in reporter constructs that had either the wildtype or the mutant EWS16T tract (6U deletions) cloned downstream of luciferase. None of the siRNAs influenced the poly(A) site usage in the wildtype constructs in which only the shorter isoform was expressed (Figure 4b). When the 6U deletion construct was used, however, the knockdown of SFPQ and hnRNPC strongly influenced poly(A) site selection. Specifically, knockdown of SFPQ promoted proximal site usage, while the knockdown of hnRNPC led to increased expression of the longer 3'UTR isoform (Figure 4c). This result was consistent with the observed binding pattern of the RBPs. hnRNPC binding was most prominent immediately downstream of the proximal poly(A) site (our unpublished PAR-CLIP data on hnRNPC), while SFPQ most likely bound to the truncated U-tract immediately upstream of the distal poly(A) site, as inferred from the pull down experiments (Figure 3a). Interestingly, knockdown of SFPQ did not affect poly(A) site selection in wildtype constructs, consistent with our earlier findings that SFPQ specifically associated with the construct carrying the 6U deletion. Thus, our results indicate that hnRNPC and SFPQ have antagonistic activity on the processing of mutant *EWSR1* pre-mRNA with hnRNPC promoting the generation of the shorter and SFPQ of the longer 3'UTR isoform.

Distal poly(A) site usage is associated with decreased *EWSR1* expression.

To determine if the choice in polyadenylation site may influence *EWSR1* expression, we performed luciferase assays on the constructs that carried the wildtype or various EWS16T deletion variants. The results shown in Supplementary Fig. 6 indicate a significant downregulation (up to 30%) of protein levels associated with EWS16T tract deletions. We attempted to corroborate our observations *in vivo*, assessing *EWSR1* mRNA expression in a set of 8 LS-related MMR-deficient and 5 sporadic MMR-proficient CRCs relative to matched, tumor-free mucosa by qPCR. Four (50%) out of the 8 LS, but only 1 (20%) out of the 5 sporadic CRC showed significantly (>1.2 fold) increased expression of the longer 3'UTR isoform (Figure 5A). The total *EWSR1* mRNA levels, however, were significantly reduced (>1.2 fold) in both LS-related and sporadic CRC tissues compared to the matched, tumor-free mucosa (Figure 5B), pointing to additional mechanisms that regulate *EWSR1* expression in

sporadic CRCs. Alternatively, qPCR interpretation may have been compromised by the presence of non-cancer cells, especially in MMR-deficient cancer specimens.

To determine the consequences of altered *EWSR1* mRNA expression *in vivo*, we performed immunohistochemical analysis (IHC) of 10 LS-related, MMR-deficient and 9 sporadic, MMR-proficient CRCs. Consistent with the data obtained at the mRNA level, the cancers displayed on average an approx. 30% reduction in EWS expression when compared to matched, tumor-free mucosa. Surprisingly however, MMR-deficient and -proficient cancers significantly differed with regard to the subcellular localization of EWS ($P < 0.001$). Tumor-free colon mucosa and adenomas from LS and sporadic CRC patients as well as sporadic adenocarcinomas showed exclusively nuclear expression (Fig. 6). In contrast, LS-related CRCs displayed diffuse cytoplasmic EWS expression (Fig. 6). These results were subsequently confirmed by IHC analysis of a tissue microarray (TMA) containing 64 sporadic and 94 LS-related CRCs: we observed an approx. 30% reduction in EWS expression in both groups, but only the LS-related cancers showed diffuse cytoplasmic staining for EWS (61% vs 3% of the sporadic cancers, $p < 0.001$). Thus, MMR-deficient CRCs, all carrying somatic EWS16T tract alterations, display a distinct subcellular EWS distribution pattern *in vivo*. Further studies, ideally performed on EWS16T-mutated CRC cell lines, are needed to assess if this is directly related to EWS16T tract alterations or, rather, an indirect consequence of MSI-associated genetic instability affecting regulators of EWSR1 protein localization such as the methyl-transferases like PRMT1, which is known to regulate the localization of EWS(21) by methylating Glycine/Arginine-rich motifs located in the arginine-glycine-glycine domains of EWS.

Discussion

EWS16T, a polyT/U tract in the 3'UTR of the *EWSR1* gene, represents a novel, quasi-monomorphic MSI target gene locus that identifies both hereditary and sporadic MMR-deficient colorectal cancers with perfect (100%) diagnostic sensitivity and specificity. Given this high diagnostic accuracy and quasi-monomorphic nature ascertained in more than 300 cancers, EWS16T has the potential to substantially facilitate and improve the accuracy of MSI detection in routine daily practice and prospective studies are now needed to determine whether EWS16T alone could replace the set of multiple diagnostic markers currently employed for MSI testing.

3' untranslated regions (3'UTRs) play an essential role in regulating the stability, subcellular localization, and translation of corresponding mRNAs via sequence-specific interactions with trans-acting factors including small RNAs and proteins(22). Several 3'UTR point mutations have been linked to the risk of developing cancer in humans(23-28). Recent reports have shown that contractions at 3'UTR mononucleotide repeats can exert deleterious effects by destabilizing the respective mRNAs associated with them and thereby be of direct pathophysiological relevance(29, 30). Characterization of 3'UTR MSI target genes thus can provide important insights into the consequences of MMR for gene expression. Our investigations of the *in vitro* and *in vivo* effects of EWS16T alterations conclusively show that contractions at this locus indeed affect multiple regulatory mechanisms such alternative polyadenylation, mRNA / protein expression and possibly subcellular localization. Given that EWS is an RNA-/DNA-binding protein with many RNA/DNA targets, these alterations provide new leads into the mechanisms behind MSI-associated colorectal tumorigenesis.

The physiological role of EWS is largely unknown but based on its domain structure the protein is thought to be involved in such diverse processes as gene expression, RNA processing / transport and cell signalling. Knockout of EWS in mice results in high postnatal lethality, defects in pre-B cell development, meiotic arrest/germ cell apoptosis, premature cellular senescence, and hypersensitivity to ionizing radiation (IR) (31). These observations suggest roles for EWS in homologous recombination, DNA damage response, and maintenance of genome integrity (32). Indeed, we recently found that *EWSR1* binds RNAs that originate at intrinsically unstable genomic loci and its knockdown increases the frequency

of double-strand DNA breaks (Kishore et al., data not shown).

With respect to tumorigenesis, genetic alterations in *EWSR1* were first observed in Ewing sarcoma, the second most common malignant bone tumor in adolescents and young adults after osteosarcoma(16, 33, 34). In about 85% of cases, an *EWSR1-FLI1* fusion protein is observed which has retained the N-terminal transcription activation domain but lost the RNA-binding domains, which are replaced with the DNA binding domain of the fusion partner (35). The fusion protein is constitutively active and has been shown to alter the transcription of several downstream targets. Ewing Sarcoma is thus largely thought of as a gain of function phenotype. Loss of the *EWSR1* function has been largely overlooked, in spite of the fact that the protein has a canonical RNA binding domain and has been shown to regulate several RNA processing events in the nucleus (36, 37). As a member of the TET (TLS/FUS, EWS, and TAF15) family of RNA- and DNA-binding proteins it has been involved in transcriptional regulation and RNA processing (38) (39) (40) (41, 42). Consistent with this potential dual role, EWS has been shown to regulate cyclin D1 transcripts both transcriptionally and at the level of splicing, with the oncogenic fusion protein EWS-FLI1 promoting the expression of the oncogenic cyclin D1b splice variant in Ewing sarcoma cells (43). More recently, EWS has been shown to regulate alternative splicing (AS) of the p53 repressor *MDM2* and to be part of the microprocessor complex that mediates the genesis of microRNAs (32, 42) (44).

In summary, the EWS16T locus represents a novel, quasi-monomorphic MSI target locus to accurately identify both hereditary and sporadic MMR deficient cancers. Contractions therein affect *EWSR1* at multiple levels including alternative polyadenylation, mRNA / protein expression and possibly subcellular localization. Our results thereby implicate the RNA-/DNA-binding protein EWSR1, critical for the maintenance of genome integrity(45), in MSI-associated colorectal tumorigenesis.

References

1. Hampel H, de la Chapelle A. The search for unaffected individuals with Lynch syndrome: do the ends justify the means? *Cancer Prev Res*. 2011;4:1-5.
2. Win AK, Lindor NM, Young JP, Macrae FA, Young GP, Williamson E, et al. Risks of primary extracolonic cancers following colorectal cancer in lynch syndrome. *J Natl Cancer Inst*. 2012;104:1363-72.
3. Boland CR, Goel A. Microsatellite instability in colorectal cancer. *Gastroenterology*. 2010;138:2073-87 e3.
4. Popat S, Hubner R, Houlston RS. Systematic review of microsatellite instability and colorectal cancer prognosis. *J Clin Oncol*. 2005;23:609-18.
5. di Pietro M, Sabates Bellver J, Menigatti M, Bannwart F, Schnider A, Russell A, et al. Defective DNA mismatch repair determines a characteristic transcriptional profile in proximal colon cancers. *Gastroenterology*. 2005;129:1047-59.
6. Duval A, Hamelin R. Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability. *Cancer Res*. 2002;62:2447-54.
7. Mendell JT, Dietz HC. When the message goes awry: disease-producing mutations that influence mRNA content and performance. *Cell*. 2001;107:411-4.
8. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Ruschoff J, et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst*. 2004;96:261-8.
9. de Kok JB, Roelofs RW, Giesendorf BA, Pennings JL, Waas ET, Feuth T, et al. Normalization of gene expression measurements in tumor tissues: comparison of 13 endogenous control genes. *Lab Invest*. 2005;85:154-9.
10. Dydensborg AB, Herring E, Auclair J, Tremblay E, Beaulieu JF. Normalizing genes for quantitative RT-PCR in differentiating human intestinal epithelial cells and adenocarcinomas of the colon. *Am J Physiol Gastrointest Liver Physiol*. 2006;290:G1067-74.
11. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*. 2001;25:402-8.
12. Zlobec I, Molinari F, Kovac M, Bihl MP, Altermatt HJ, Diebold J, et al. Prognostic and predictive value of TOPK stratified by KRAS and BRAF gene alterations in sporadic, hereditary and metastatic colorectal cancer patients. *Br J Cancer*. 2010;102:151-61.
13. Lugli A, Tzankov A, Zlobec I, Terracciano LM. Differential diagnostic and functional role of the multi-marker phenotype CDX2/CK20/CK7 in colorectal cancer stratified by mismatch repair status. *Mod Pathol*. 2008;21:1403-12.
14. Allred DC, Harvey JM, Berardo M, Clark GM. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol*. 1998;11:155-68.
15. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature*. 2008;452:872-6.

16. Ludwig JA. Ewing sarcoma: historical perspectives, current state-of-the-art, and opportunities for targeted therapy in the future. *Curr Opin Oncol*. 2008;20:412-8.
17. Mayr C, Bartel DP. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*. 2009;138:673-84.
18. Martin G, Gruber AR, Keller W, Zavolan M. Genome-wide analysis of pre-mRNA 3' end processing reveals a decisive role of human cleavage factor I in the regulation of 3' UTR length. *Cell Rep*. 2012;1:753-63.
19. Kuwano Y, Pullmann R, Jr., Marasa BS, Abdelmohsen K, Lee EK, Yang X, et al. NF90 selectively represses the translation of target mRNAs bearing an AU-rich signature motif. *Nucleic Acids Res*. 2010;38:225-38.
20. Lutz CS, Cooke C, O'Connor JP, Kobayashi R, Alwine JC. The snRNP-free U1A (SF-A) complex(es): identification of the largest subunit as PSF, the polypyrimidine-tract binding protein-associated splicing factor. *RNA*. 1998;4:1493-9.
21. Araya N, Hiraga H, Kako K, Arao Y, Kato S, Fukamizu A. Transcriptional down-regulation through nuclear exclusion of EWS methylated by PRMT1. *Biochem Biophys Res Commun*. 2005;329:653-60.
22. Lopez de Silanes I, Quesada MP, Esteller M. Aberrant regulation of messenger RNA 3'-untranslated region in human cancer. *Cell Oncol*. 2007;29:1-17.
23. Chatterjee S, Pal JK. Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. *Biol Cell*. 2009;101:251-62.
24. Wiestner A, Tehrani M, Chiorazzi M, Wright G, Gibellini F, Nakayama K, et al. Point mutations and genomic deletions in CCND1 create stable truncated cyclin D1 mRNAs that are associated with increased proliferation rate and shorter survival. *Blood*. 2007;109:4599-606.
25. Heinimann K, Thompson A, Locher A, Furlanetto T, Bader E, Wolf A, et al. Nontruncating apc germ-line mutations and mismatch repair deficiency play a minor role in apc mutation-negative polyposis. *Cancer Res*. 2001;61:7616-22.
26. Pietschmann A, Mehdipour P, Mehdipour P, Atri M, Hofmann W, Hosseini-Asl SS, et al. Mutation analysis of BRCA1 and BRCA2 genes in Iranian high risk breast cancer families. *J Cancer Res Clin Oncol*. 2005;131:552-8.
27. Mao G, Pan X, Gu L. Evidence that a mutation in the MLH1 3'-untranslated region confers a mutator phenotype and mismatch repair deficiency in patients with relapsed leukemia. *J Biol Chem*. 2008;283:3211-6.
28. Paun BC, Cheng Y, Leggett BA, Young J, Meltzer SJ, Mori Y. Screening for microsatellite instability identifies frequent 3'-untranslated region mutation of the RB1-inducible coiled-coil 1 gene in colon tumors. *PLoS One*. 2009;4:e7715.
29. Shin J, Yuan Z, Fordyce K, Sreeramoju P, Kent TS, Kim J, et al. A del T poly T (8) mutation in the 3' untranslated region (UTR) of the CDK2-AP1 gene is functionally significant causing decreased mRNA stability resulting in decreased CDK2-AP1 expression in human microsatellite unstable (MSI) colorectal cancer (CRC). *Surgery*. 2007;142:222-7.

30. Ruggiero T, Olivero M, Follenzi A, Naldini L, Calogero R, Di Renzo MF. Deletion in a (T)₈ microsatellite abrogates expression regulation by 3'-UTR. *Nucleic Acids Res.* 2003;31:6561-9.
31. Li H, Watford W, Li C, Parmelee A, Bryant MA, Deng C, et al. Ewing sarcoma gene EWS is essential for meiosis and B lymphocyte development. *J Clin Invest.* 2007;117:1314-23.
32. Paronetto MP, Minana B, Valcarcel J. The Ewing sarcoma protein regulates DNA damage-induced alternative splicing. *Mol cell.* 2011;43:353-68.
33. Delattre O, Zucman J, Plougastel B, Desmaze C, Melot T, Peter M, et al. Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature.* 1992;359:162-5.
34. Zucman J, Delattre O, Desmaze C, Plougastel B, Joubert I, Melot T, et al. Cloning and characterization of the Ewing's sarcoma and peripheral neuroepithelioma t(11;22) translocation breakpoints. *Genes Chromosom Cancer.* 1992;5:271-7.
35. Sankar S, Lessnick SL. Promiscuous partnerships in Ewing's sarcoma. *Cancer Genet.* 2011;204:351-65.
36. Petermann R, Mossier BM, Aryee DN, Khazak V, Golemis EA, Kovar H. Oncogenic EWS-Fli1 interacts with hRPB7, a subunit of human RNA polymerase II. *Oncogene.* 1998;17:603-10.
37. Zinszner H, Albalat R, Ron D. A novel effector domain from the RNA-binding protein TLS or EWS is required for oncogenic transformation by CHOP. *Genes Dev.* 1994;8:2513-26.
38. Bertolotti A, Melot T, Acker J, Vigneron M, Delattre O, Tora L. EWS, but not EWS-FLI-1, is associated with both TFIID and RNA polymerase II: interactions between two members of the TET family, EWS and hTAFII68, and subunits of TFIID and RNA polymerase II complexes. *Mol Cell Biol.* 1998;18:1489-97.
39. Knoop LL, Baker SJ. The splicing factor U1C represses EWS/FLI-mediated transactivation. *J Biol Chem.* 2000;275:24865-71.
40. Zhang D, Paley AJ, Childs G. The transcriptional repressor ZFM1 interacts with and modulates the ability of EWS to activate transcription. *J Biol Chem.* 1998;273:18086-91.
41. Chansky HA, Hu M, Hickstein DD, Yang L. Oncogenic TLS/ERG and EWS/Fli-1 fusion proteins inhibit RNA splicing mediated by YB-1 protein. *Cancer Res.* 2001;61:3586-90.
42. Dutertre M, Sanchez G, De Cian MC, Barbier J, Dardenne E, Gratadou L, et al. Cotranscriptional exon skipping in the genotoxic stress response. *Nat Struct Mol Biol.* 2010;17:1358-66.
43. Sanchez G, Delattre O, Auboeuf D, Dutertre M. Coupled alteration of transcription and splicing by a single oncogene: boosting the effect on cyclin D1 activity. *Cell Cycle.* 2008;7:2299-305.
44. Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, et al. The Microprocessor complex mediates the genesis of microRNAs. *Nature.* 2004;432:235-40.
45. Paronetto MP, Minana B, Valcarcel J. The Ewing sarcoma protein regulates DNA damage-induced alternative splicing. *Mol Cell.* 2011;43:353-68.

ACKNOWLEDGEMENTS

We thank the patients and their families for participating in this study and the treating physicians for providing pertinent clinical information and tumor specimens. In particular, we are indebted to Prof. G. Spagnoli, Dept. Biomedicine, University of Basel, for sharing his expertise on cell culturing and harvesting, with Dr. Luca Quagliata and Dr. Nitish Mittal for their effort in cell culture. We also thank Michèle Attenhofer, Nemja Boesch and Sibylle Bertschin for excellent technical assistance.

Funding

The study was supported by grants from the European Molecular Biology Organization (long-term fellowship 184-2009 to S.K.), the Academy of Finland (P.P.), the Finnish Cancer Organizations (P.P.), the Sigrid Juselius Foundation (P.P.), Biocentrum Helsinki (P.P.), the European Research Council FP7-ERC-232635 (P.P.), the University of Basel (M.Z.), the Krebsliga beider Basel (K.H.), the Krebsliga Zentralschweiz (K.H.) and Oncosuisse (K.H.).

AUTHOR CONTRIBUTIONS

S.K. performed the *in vitro* transcription assays, protein pull down assays, poly(A) site selection assays and luciferase assays; S.P. and M.K. performed the MSI analyses on the Swiss patients and qPCR experiments; P.P. and A.G. performed the MSI analyses on the Finnish patients, S.P., F.T. and L.T. analyzed CRC whole tissue and microarray sections; H.A. and G.M. performed immunohistopathological analyses and F.W. FISH analyses on the extraosseous Ewing sarcoma; V.M. cultured and provided the MMR proficient and deficient colorectal cancer cell lines; S.P., S.K., M.K., M.Z. and K.H. analyzed the data and wrote the manuscript; all authors contributed and agreed to the final version of the manuscript.

Tables

Table1

Samples analyzed	n	MSI status	EWS16T tract status		
			wildtype n (%)	contraction n (%)	expansion n (%)
MMR deficient cancers	128	MSI-High	0	126 (98.4)	2 (1.6)
• Colorectal	104		0	102 (98.0)	2 (2.0)
• <i>MLH1</i> deficient	76		0	76 (100)	0
• <i>MSH2</i> deficient	28		0	26 (92.9)	2 (7.1)
• Gastric	17		0	17 (100)	0
• <i>MLH1</i> deficient	16		0	16 (100)	0
• <i>MSH2</i> deficient	1		0	11 (100)	0
• Endometrial	7		0	7 (100)	0
• <i>MLH1</i> deficient	4		0	4 (100)	0
• <i>MSH2</i> deficient	3		0	3 (100)	0
MMR proficient cancers	191	MSS/ MSI-Low	190 (100)	0	0
• Colorectal	124	MSS	124 (100)	0	0
• Colorectal	12	MSI-Low	12 (100)	0	0
• Gastric	55	MSS	55 (100)	0	0
MMR proficient colon adenomas	37	MSS	37 (100)	0	0
• Lynch syndrome-related	10		10 (100)	0	0
• <i>MLH1</i> deficient	8		8 (100)	0	0
• <i>MSH2</i> deficient	2		2 (100)	0	0
• Sporadic	27		27 (100)	0	0

Table 1: EWS16T tract instability in 34 colon adenoma and 319 cancer samples.

Microsatellite instability was determined according to Umar et al. (8). MMR: mismatch repair;

MSI: microsatellite instability; MSS: microsatellite stable.

FIGURE LEGENDS

Figure 1: EWS16T tract instability in MMR deficient cancers as determined by capillary electrophoresis of fluorescently labeled PCR products. The dotted line corresponds to the wild-type allele (16T). MMR-proficient cancers **(a)** colorectal, **(b)** gastric; MMR-deficient cancers with expansion **(c)** or contraction of EWS16T **(d-h)**: **(c)** colorectal (*MSH2* germline mutation), **(d)** colorectal (*MSH2*), **(e)** colorectal (*MLH1* promoter hypermethylation), and cancers with an *MLH1* germline mutation **(f-h)**: **(f)** colorectal, **(g)** gastric, **(h)** endometrial.

Figure 2: Poly(A) site selection assay through multiplexed PCR. **a)** Schematic diagram of the multiplexed semi-quantitative protocol for poly(A) site selection assay. A single forward primer, which was either vector specific (for analyzing poly(A) usage of reporter constructs) or EWSR1 -3'UTR specific (for analyzing endogenous EWSR1 poly(A) usage), was used along with two terminally anchored reverse primers to yield amplicons representing both long and short variants of EWSR1 3'UTR in a single PCR reaction. **b)** Poly(A) site usage in EWSR1 3'UTR constructs with variable poly(T/U) tracts lengths cloned in psiCHECK-2-SPAm vector along with the quantification of distal poly(A) site usage (percentages). **c)** Poly(A) site usage in endogenous EWSR1 gene across various MMR-deficient and -proficient cell lines.

Figure 3: Identification of proteins associated with the 3'UTR of *EWSR1* in a pull down assay. **A)** Colloidal blue stained SDS PAGE gel showing proteins specifically interacting with *in vitro* transcribed, S1 aptamer-tagged *EWSR1*-wt or -6del 3'UTR. "No RNA" refers to the control sample where beads were incubated with cellular lysate without a prior incubation with RNA, allowing detection of any proteins interacting non-specifically with the MyOne Streptavidin beads. Bands specifically present in either wt or 6del lanes were excised and submitted to protein identification by mass spectrometry. Identified proteins have been labeled next to the corresponding bands. **B)** Western blot confirming the results of mass spectrometry. Western blot with antibody raised against HuR **(a)**, SFPQ **(b)** and NF90 **(c)** proteins. The membrane from **(b)** was used to reblot with anti-NF90 in **(c)**. Asterix in **(c)** depicts the background signal from SFPQ in **(b)**.

Figure 4: Knockdown of RNA binding proteins identified from pull down assays showing that SFPQ and hnRNPC specifically modulate poly(A) site usage in *EWSR1* 3'UTR reporter constructs. **a)** Western blot showing the knock down of the RNA-binding proteins after 72 hours. Poly(A) site usage in **b)** *EWSR1*-wt-psiCHECK-2-SPAm and **c)** *EWSR1*-6del-psiCHECK-2-SPAm reporter constructs upon knock down of RNA binding proteins. **d)** Quantification of the relative poly(A) site usage in **c)**.

Figure 5: Quantitative Real-Time PCR showing *EWSR1* expression in tumor relative to matched normal tissue. **A)** Difference in poly(A) site usage in endogenous *EWSR1* mRNA (distal/proximal) and **B)** difference in total *EWSR1* mRNA levels. Samples showing more than 1.2 fold difference are indicated with an asterix.

Figure 6: Immunohistochemical staining of EWS. **a)** Normal colonic mucosa from sporadic CRC patient. **b)** Sporadic, MMR proficient colon adenoma with nuclear expression. **c)** Sporadic, MMR proficient colon adenocarcinoma with nuclear expression. **d)** Normal colonic mucosa from Lynch syndrome-related CRC patient. **e)** MMR deficient colon adenoma with nuclear expression. **f)** MMR deficient colon adenocarcinoma with nuclear and diffuse cytoplasmic expression.

CRC, colorectal cancer.

Figure 1

Mismatch Repair

Proficient cancers

Deficient cancers

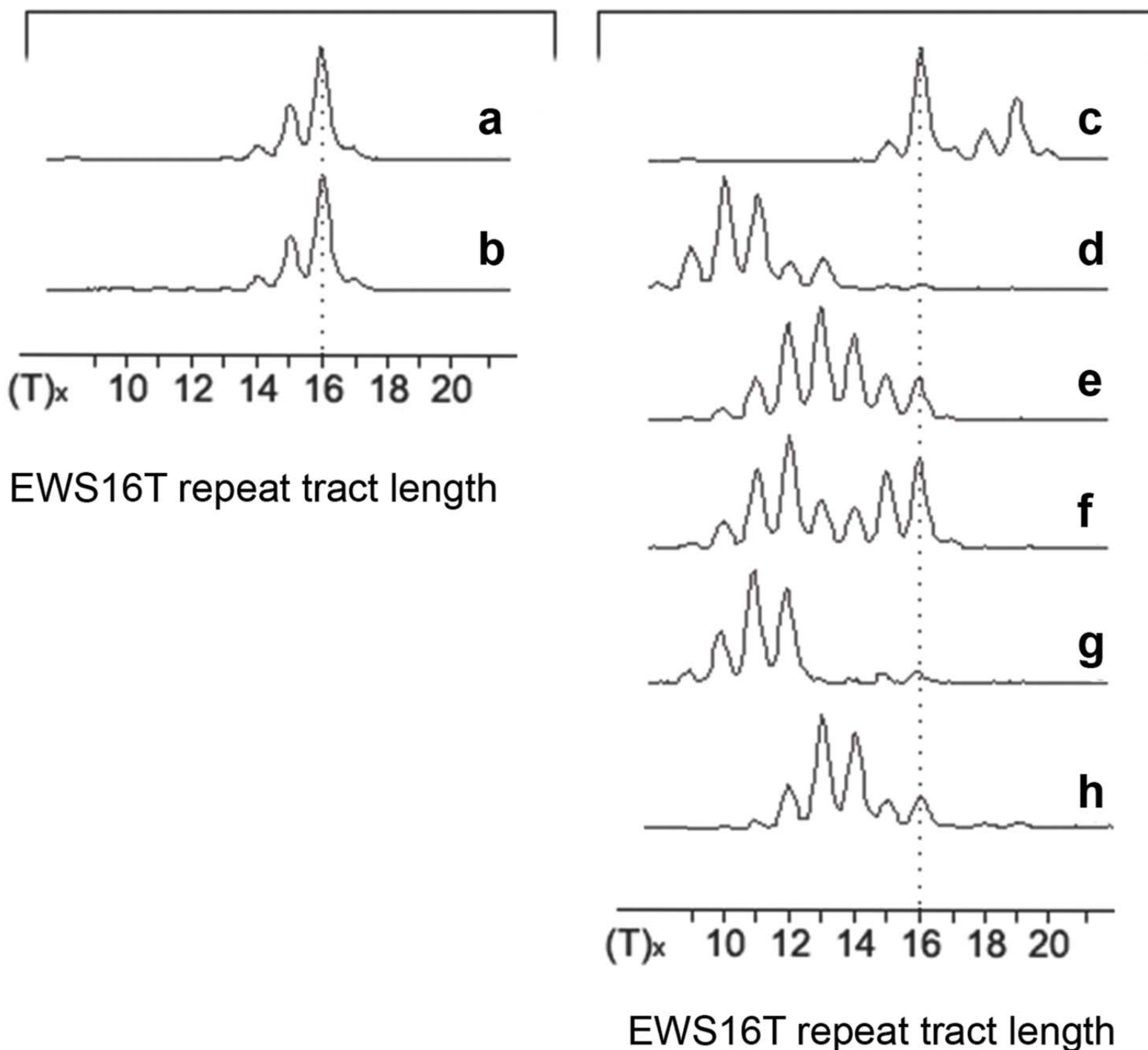


Figure 2

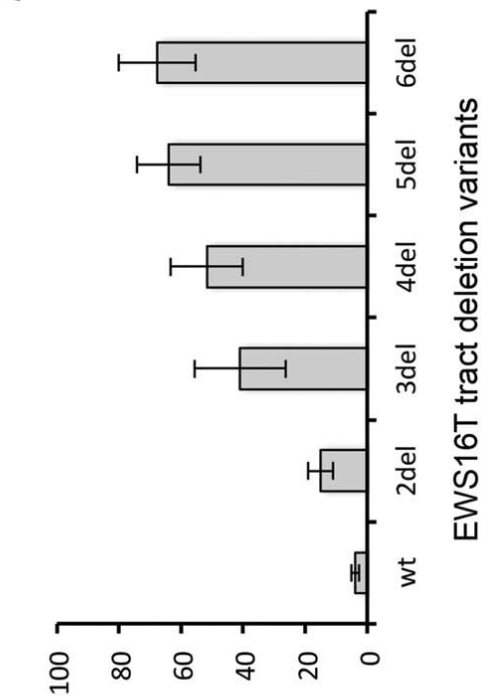
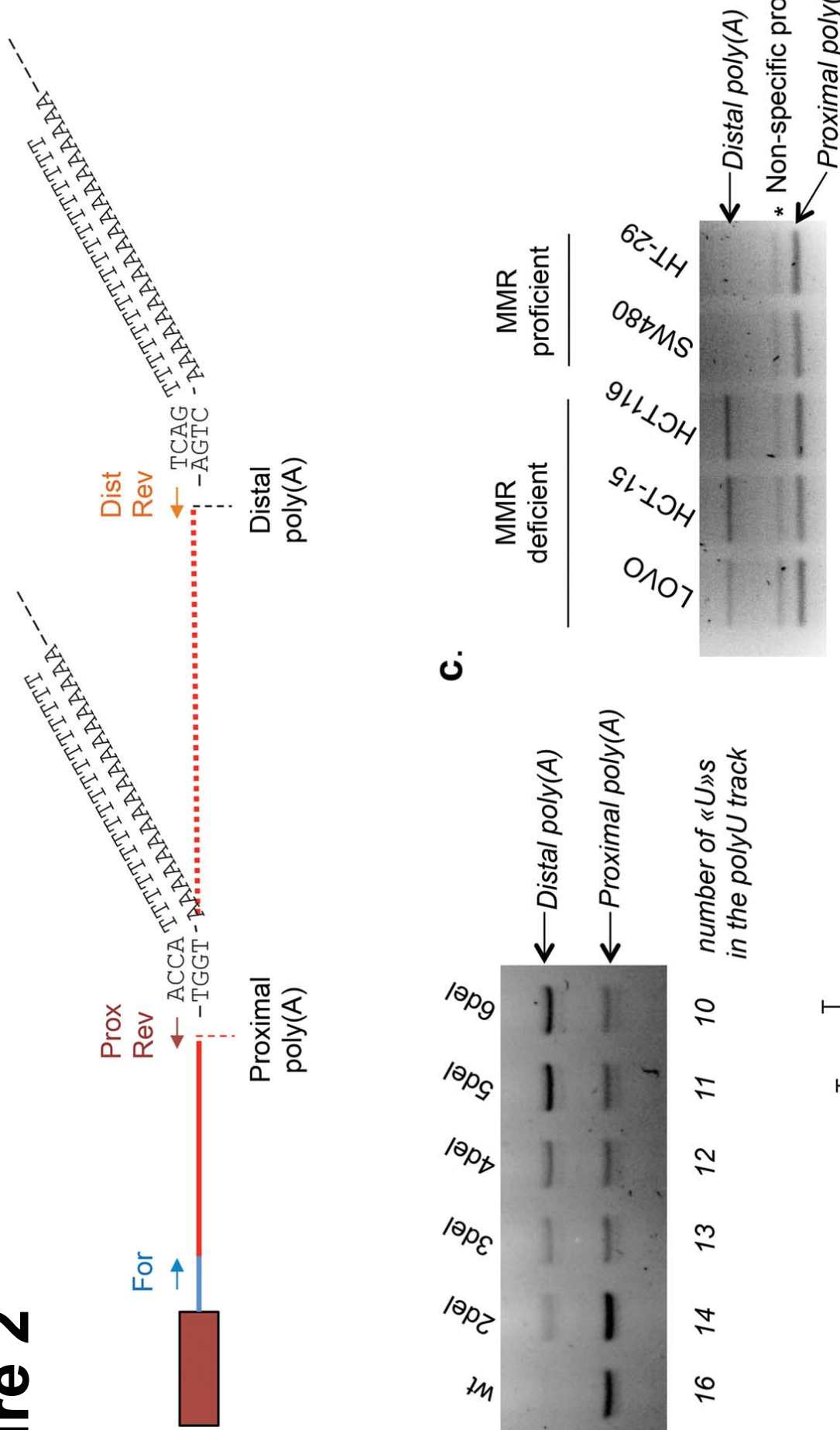


Figure 3A

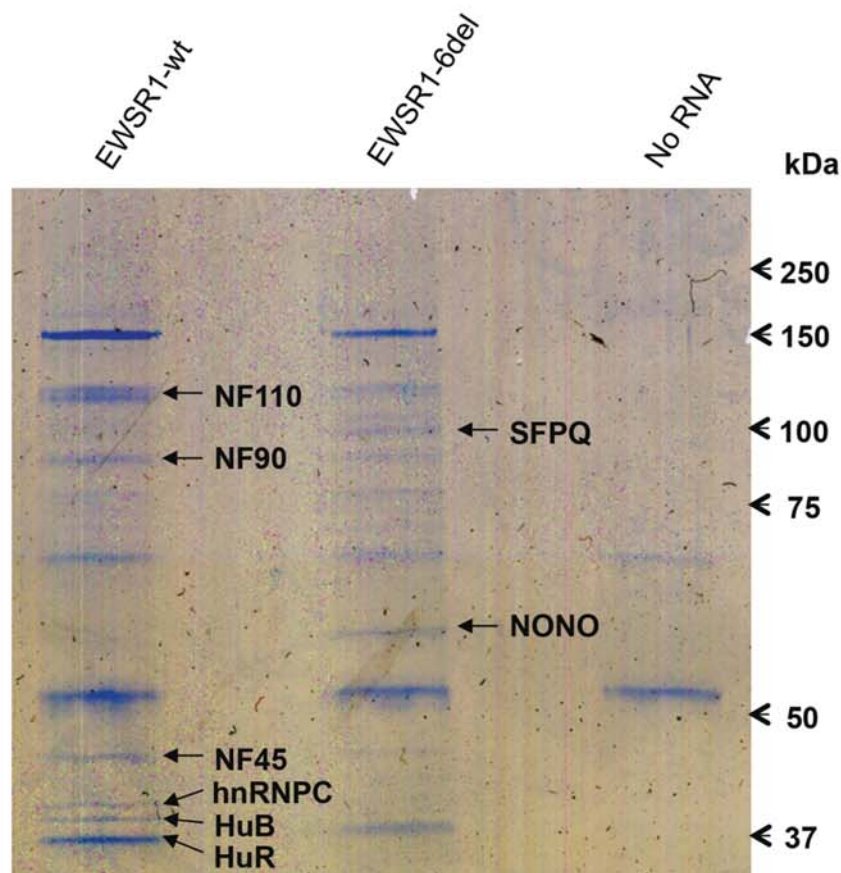


Figure 3B

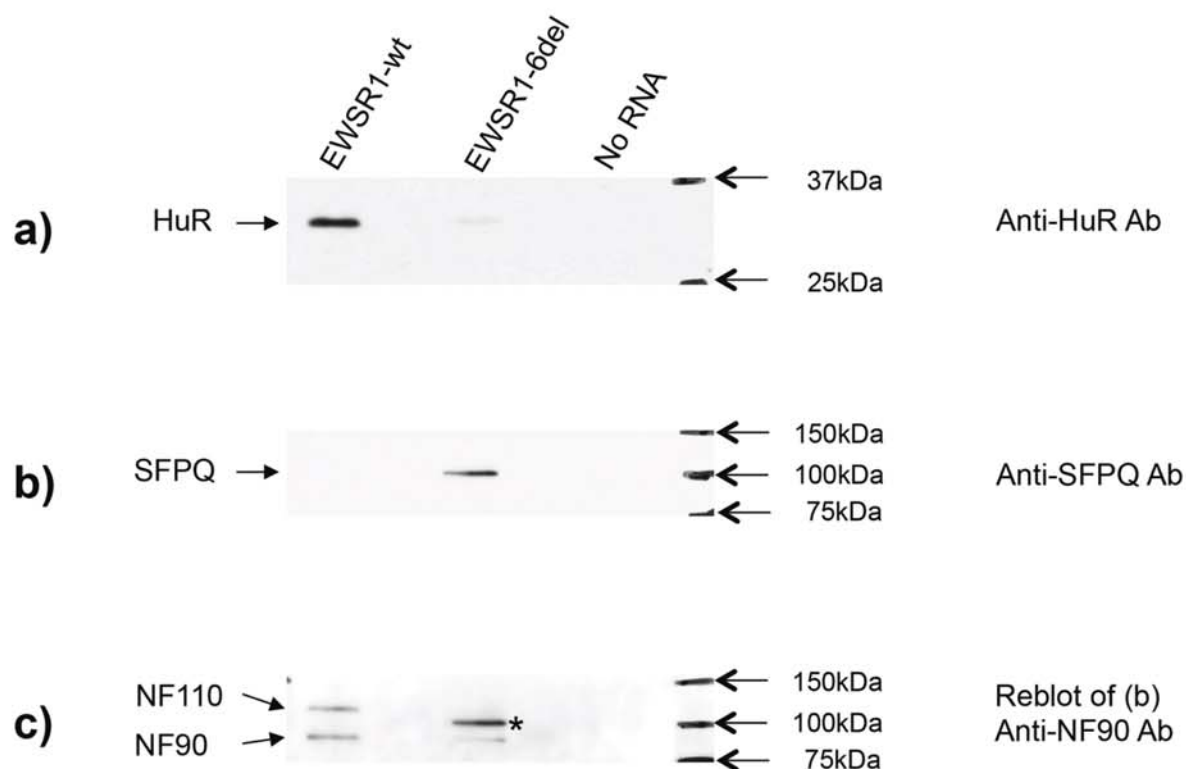


Figure 4

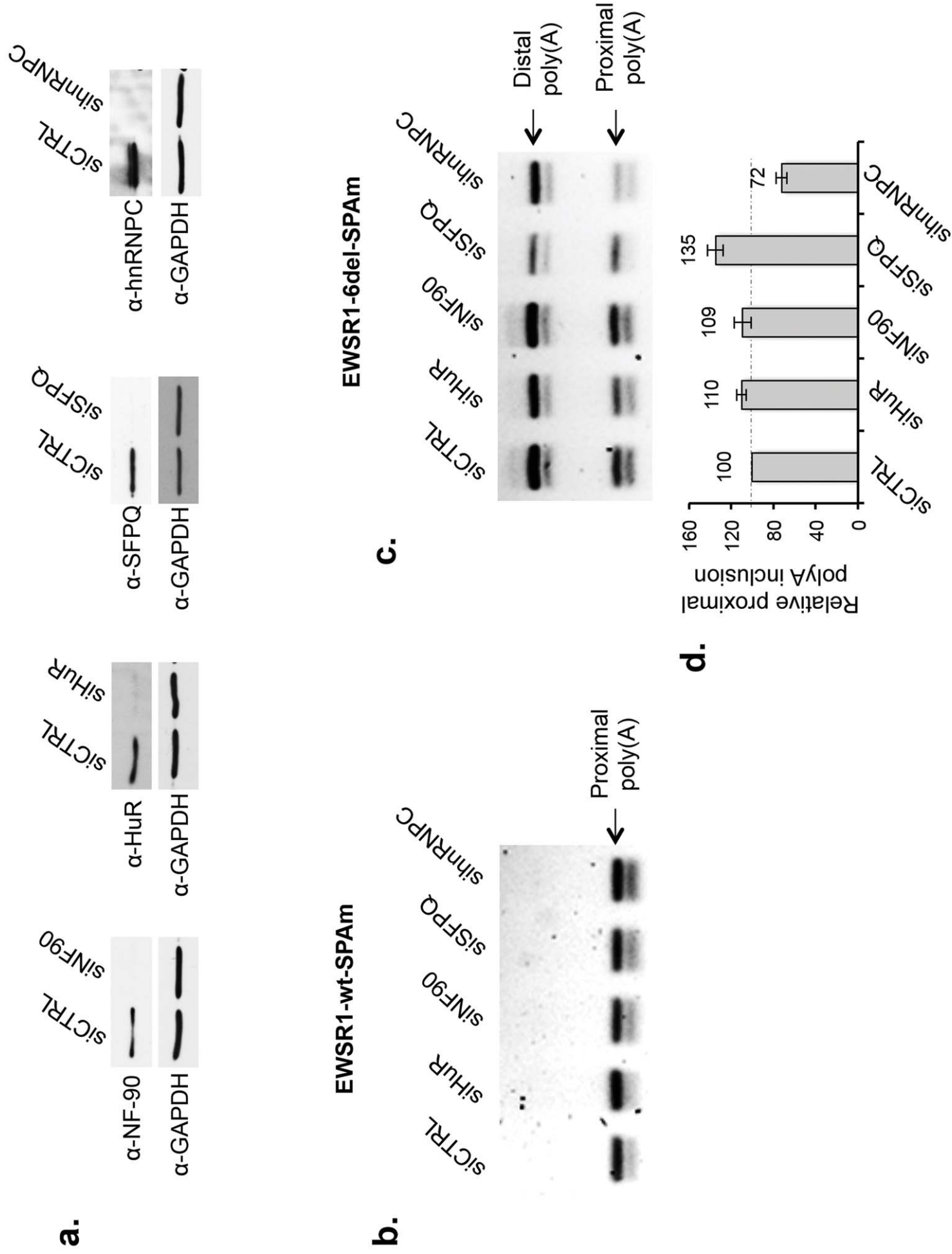


Figure 5A

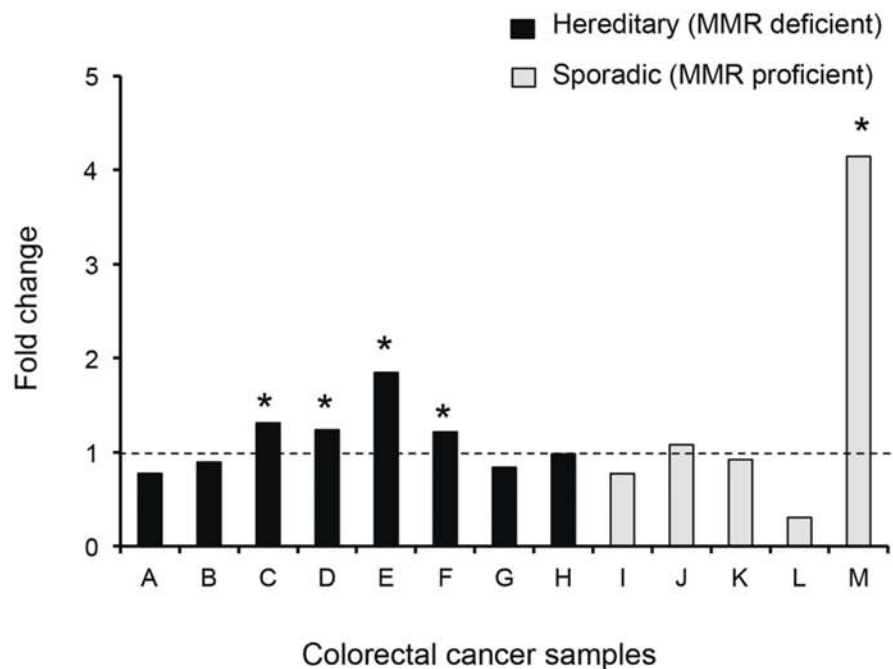


Figure 5B

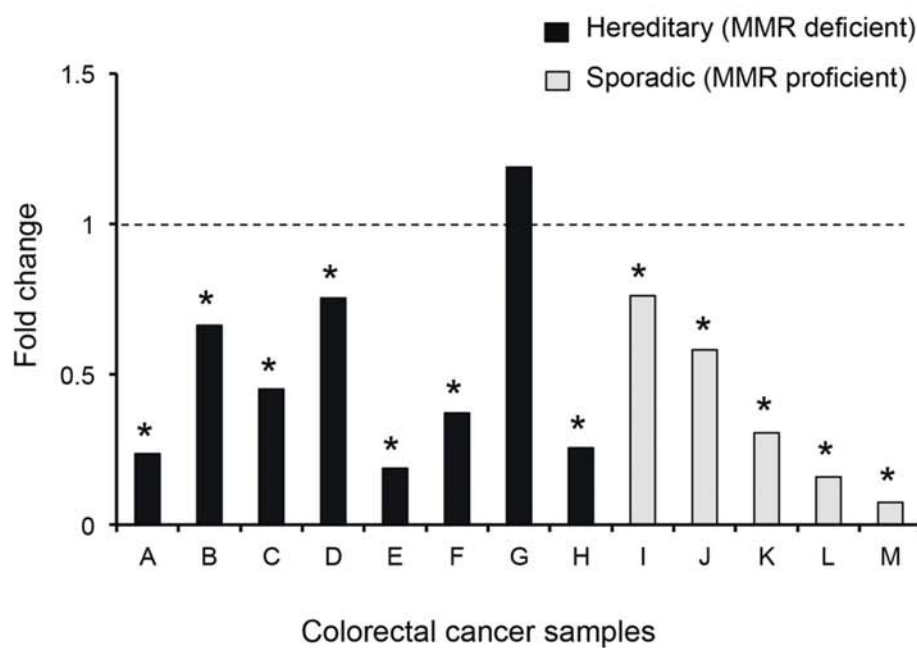


Figure 6

